Winter 2008

CS 790-02: Advanced Data Mining

Guozhu Dong
Wright State University - Main Campus, guozhu.dong@wright.edu

CS 790-02: Advanced Data Mining

Winter, 2008
Syllabus

Description: This course covers advanced topics in data mining. The topics can be roughly classified along the following dimensions: pattern/model types, mining/analysis techniques, and data types. The course will examine sequence mining, text mining, graph mining, data cube mining, microarray gene expression mining, stream mining, time series mining, web mining, bioinformatics, privacy issues, etc.

Text books:

- Data Mining: Concepts and Techniques, 2nd edition. J. Han and M. Kamber. Morgan Kaufmann.
- Sequence Data Mining, Guozhu Dong and Jian Pei, Springer, 2007.

Prerequisite: CS705 (Introduction to data mining), or consent of the instructor.
Instructor: Dr. Guozhu Dong.
Office: Joshi 383
Phone No.: (937)-775-5066
Email: guozhu.dong@wright.edu
Class details:

- Room: Joshi 384
- Time: TBA (?6:05-7:20 TTh?)

Office hours: TBA. Use e-mail for short questions.
Format: This course will draw materials from the text books, and from a range of recent research papers. Students will need to submit review reports on the papers/chapters, actively participate in class discussions, present papers, complete a programming project, and do two exams.
Problem brainstorming: Each student needs to identify, motivate, formalize, and propose potential solutions of, a problem (related to the course materials) that the student wishes to solve. Students are not required to completely solve these problems, but they are encouraged to do so.
Term Project: Students will complete an implementation project, on instructor-approved topics (mostly based on published research papers). There will be a demo and a report, and the codes must be submitted. Students should talk to the instructor early to select paper for the project.
Evaluation: Final grades in the course will be determined as follows:

- Paper/chapter reviews: 10%
- Paper/chapter presentations: 20%
- Class participation: 10%
- Problem brainstorming: 5%
- Term project: 35% (Report: 20%; Demo: 15%)
- Exams: 20%.

Resources:

- http://www.kdnuggets.com/
- www.scholar.google.com
- DBLP Bibliography Server: http://www.informatik.uni-trier.de/~ley/db/index.html
- ACM Digital Library: http://www.acm.org/dl/ (Free access from WRIGHT domain.)
- ResearchIndex (citeseer):http: //citeseer.ist.psu.edu/ (not up to date)